



# On the ASTA Property in a Feedback Processor-Sharing Queue

Isi Mitrani, Philippe Robert

## ► To cite this version:

Isi Mitrani, Philippe Robert. On the ASTA Property in a Feedback Processor-Sharing Queue. [Research Report] RR-5423, INRIA. 2004, pp.7. inria-00070583

**HAL Id: inria-00070583**

**<https://inria.hal.science/inria-00070583>**

Submitted on 19 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *On the ASTA Property in a Feedback Processor-Sharing Queue*

Isi Mitrani — Philippe Robert

N° 5423

Déembre 2004

Thème COM

A large blue rectangle occupies the lower half of the page. Overlaid on it is a large, light gray stylized 'R' logo. To the right of the 'R', the words 'Rapport de recherche' are written in a white serif font. A horizontal gray brushstroke is positioned below the text.

*Rapport  
de recherche*





## On the ASTA Property in a Feedback Processor-Sharing Queue

Isi Mitrani<sup>\*</sup>, Philippe Robert<sup>†</sup>

Thème COM — Systèmes communicants  
Projet Rap

Rapport de recherche n° 5423 — Deembre 2004 — 7 pages

**Abstract:** A stationary M/M/1 Processor-Sharing queue with Bernoulli feedbacks is considered in this paper. Jobs arrive externally with rate  $\lambda$  and require services whose average is  $1/\mu$ ; after completing a service, a job leaves the system with probability  $\alpha$  and rejoins the queue with probability  $1 - \alpha$ . The queue is stable when the offered load,  $\rho = \lambda/(\alpha\mu)$ , satisfies  $\rho < 1$ . The distribution of the queue seen by a tagged job which has joined the queue  $n$  times already is investigated.

**Key-words:** Processor-Sharing. Feed Back. ASTA

<sup>\*</sup> School of Computing Science, University of Newcastle, NE1 7RU, UK

<sup>†</sup> RAP Project, INRIA, B.P. 105, Domaine de Voluceau, 78153 Le Chesnay, France

## La propriété ASTA des files d'attente processor-sharing avec feed-back

**Résumé :** On étudie la distribution vue par un client qui revient pour la  $n$ -ième fois dans une file processor-sharing avec bouclage.

**Mots-clés :** Partage égalitaire. ASTA.

The motivation for this short note arose from an e-mail discussion between I. Mitrani, W. Whitt and M. Reiman, concerning, essentially, a stationary M/M/1 Processor-Sharing queue with Bernoulli feedbacks. Jobs arrive externally with rate  $\lambda$  and require services whose average is  $1/\mu$ ; after completing a service, a job leaves the system with probability  $\alpha$  and rejoins the queue with probability  $1 - \alpha$ . The queue is stable when the offered load,  $\rho = \lambda/(\alpha\mu)$ , satisfies  $\rho < 1$ .

The question is whether a tagged job which has joined the queue  $n$  times already, sees the equilibrium queue size distribution upon the  $n + 1$ 'st visit or not. Two opposing arguments were put forward. On the one hand, this model is a simple special case of a separable queueing network and therefore, by the ASTA property (Arrivals See Time Averages), the queue size distribution at arrival instants, external and internal, is the same as the equilibrium distribution:

$$p_i = (1 - \rho)\rho^i, \quad i = 0, 1, \dots \quad (1)$$

However, the ASTA property in networks with probabilistic routing refers to the sequence of *all* arrival instants, not the particular subsequence consisting of the consecutive feedbacks by a given job. A priori, it is possible that such a selection of observation points affects the observed distribution.

The counter-argument proceeds as follows: suppose that the tagged job has fed back many times, and hence its total sojourn in the queue has been long. During all that time, it has occupied a fraction of the available service capacity (according to the PS scheduling policy), leaving, in effect, a slower server for the other jobs. Therefore, at the  $n + 1$ 'st feedback instant, the queue will tend to be longer than the one seen when the job first arrived into the system.

The long-term effect of a job that does not depart can be quantified. Consider a modified model where, in addition to the 'normal' jobs which come and go, there is one 'special' job which feeds back with probability 1, i.e. never leaves the system. Then, solving the Birth-and-Death equations for the number of normal jobs in the system, we find a different equilibrium distribution from (1):

$$p_i = (1 - \rho)^2(i + 1)\rho^i, \quad i = 0, 1, \dots \quad (2)$$

The average queue size corresponding to (2) is  $2\rho/(1 - \rho)$ . Thus, the presence of a long-lasting job effectively doubles the average queue size of normal jobs. However, the distribution (2) is seen by random observers and normal arrivals; a priori, the distribution seen by the special job may be different.

Although there has been a considerable amount of work on ASTA [2, 4, 5, 6]), the above question appears to be open. Most of the available results, in the context of networks with probabilistic routing, concern the sequence of all arrival instants at a node, rather than the consecutive visits to it by a particular job. The network models discussed by Kelly [2] allow a job to be characterized by the route it takes; the corresponding ASTA result (Corollary 3.5) applies to jobs that have taken a particular route. However, those models specifically exclude immediate feedbacks and so do not cover our system. A special case of a feedback queue was examined by Lam and Shankar [3], who showed that the queue size distribution at

feedback instants is independent of the number of feedbacks. However, the queueing policy in that model was FIFO. This leaves a small gap in the ASTA literature, which the present note aims to fill.

The answer to the raised question is provided by the following:

**Proposition 1** *In the open PS queue with Bernoulli feedbacks, if a tagged job sees queue size distribution (1) at its  $n$ 'th feedback instant, then it sees (1) at the  $n + 1$ 'st feedback instant (note that the job itself is not included in the number that it sees).*

Since the tagged job sees distribution (1) when it first arrives into the system (that is the PASTA property – Poisson Arrivals See Time Averages), Proposition 1 implies that it sees distribution (1) at every feedback instant, regardless of their number.

Thus, while the tagged job remains in the system, it certainly causes the queue size to increase. However, the job itself does not see that increase. This apparent paradox can be explained intuitively by remarking that when the queue is short, the tagged job circulates faster. Thus, it tends to see shorter queues more frequently than a random observer. Similarly, when the queue is long, the tagged job circulates slowly and hence tends to see longer queues less frequently than a random observer. It turns out that this observational bias exactly counteracts the lengthening effect the tagged job has on the queue.

Proposition 1 shows that also in the modified model, where one special job feeds back with probability 1, the distribution of normal jobs that it sees at feedback instants is (1), despite the fact that random observers and normal arrivals see distribution (2). In fact, we shall prove a more general result:

**Proposition 2** *If, in addition to the normal jobs, there are  $k$  special jobs which feed back with probability 1, then the stationary distribution of the number of normal jobs seen by a given special job at feedback instants is equal to the stationary distribution of the number of normal jobs in a system with  $k - 1$  special jobs.*

Proposition 2 is consistent with an ASTA property that holds in mixed queueing networks, see [6]. However, that result refers to the state seen at all consecutive service completion instants, regardless of the identity of the completing job. Here we are concerned with the state seen at a sub-sequence of those instants, i.e., the consecutive service completions for a given (arbitrary) special job. The method used to establish Proposition 2 is of some interest in its own right.

## Proof of Proposition 1

Let  $q_{i,j}$  be the conditional probability that the tagged job sees  $j$  jobs in the system at the completion of its service, given that it saw  $i$  jobs when the service started. Using the memoryless property of the exponential distribution, and conditioning upon the nature of the next event to occur (either an arrival or a service completion), we can write a set of recurrence equations:

$$q_{i,j} = \frac{\lambda}{\lambda + \mu} q_{i+1,j} + \frac{\mu}{\lambda + \mu} \left\{ \frac{i}{i+1} [\alpha q_{i-1,j} + (1 - \alpha) q_{i,j}] + \frac{1}{i+1} \delta_{i,j} \right\}, \quad (3)$$

where  $\delta_{i,j} = 1$  if  $i = j$  and 0 otherwise. The first term in the right-hand side of (3) expresses the fact that, if the next event is an arrival, then the tagged job may be considered to start its service in the presence of  $i + 1$  other jobs. The second term covers the case when the next event is a service completion: if one of the other jobs completes, then the tagged job sees  $i - 1$  or  $i$  jobs depending on whether there is a departure or a feedback; if the tagged job completes, then the state seen is  $j$  iff  $i = j$ .

Multiplying both sides of (3) by  $(\lambda + \mu)(i + 1)$  and rearranging, that equation can be rewritten as

$$q_{i,j}[i\alpha(\rho + 1) + \alpha\rho + 1] = (i + 1)\alpha\rho q_{i+1,j} + i\alpha q_{i-1,j} + \delta_{i,j} . \quad (4)$$

It is convenient to introduce the generating functions

$$g_j(z) = \sum_{i=0}^{\infty} q_{i,j} z^i , \quad j = 0, 1, \dots . \quad (5)$$

If the tagged job saw distribution (1) at its  $n$ 'th feedback, then, according to the definitions of  $q_{i,j}$  and  $g_j(z)$ , the probability that it sees  $j$  jobs at the  $n + 1$ 'st feedback is given by

$$q_{\cdot,j} = (1 - \rho)g_j(\rho) , \quad j = 0, 1, \dots . \quad (6)$$

Multiplying (4) by  $z^i$  and summing over all  $i$ , those recurrences can be transformed into first order differential equations:

$$\alpha[\rho - (\rho + 1)z + z^2]g'_j(z) = [1 + \alpha\rho - \alpha z]g_j(z) - z^j , \quad j = 0, 1, \dots . \quad (7)$$

Now, setting  $z = \rho$  in (7), the left-hand side vanishes and we obtain

$$g_j(\rho) = \rho^j ,$$

or, according to (6),

$$q_{\cdot,j} = (1 - \rho)\rho^j .$$

This establishes Proposition 1.

## Proof of Proposition 2

We use a time-change argument similar to the one in Grishechkin [1]. Let  $B = \{B(t); t \geq 0\}$  be the Birth-and-Death process which, when in state  $i$ , has a birth rate of  $(i + k)\lambda$  and death rate of  $i\alpha\mu$ . It is easily verified that  $B$  is stable when  $\rho < 1$  and its equilibrium distribution is given by

$$\pi_k(i) = \frac{(i + k - 1)!}{i!(k - 1)!} \rho^i (1 - \rho)^k , \quad i = 0, 1, \dots . \quad (8)$$

When  $k = 1$ , this coincides with (1).



Define the (strictly increasing) function

$$A(t) = \int_0^t [B(s) + k] ds , \quad (9)$$

and let  $C(t)$  be its inverse:  $C(A(t)) = A(C(t)) = t$ . Then consider the process  $L = \{L(t) = B(C(t)); t \geq 0\}$ . Since the jumps of  $L$  are uniquely mapped to jumps of  $B$ ,  $L$  is also a Birth-and-Death process. To obtain its birth rate, when in state  $i$ , we write

$$\begin{aligned} P(L(t+h) = i+1 \mid L(t) = i) &= P(B(C(t+h)) = i+1 \mid B(C(t)) = i) \\ &= P(B[C(t) + hC'(t)] = i+1 \mid B(C(t)) = i) + o(h) \\ &= (i+k)\lambda h [C'(t) \mid B(C(t)) = i] + o(h) , \end{aligned} \quad (10)$$

where, according to (9) and the definition of  $C(t)$ ,

$$[C'(t) \mid B(C(t)) = i] = \frac{1}{A'(C(t))} = \frac{1}{B(C(t)) + k} = \frac{1}{i+k} . \quad (11)$$

Substituting (11) into (10), dividing by  $h$  and letting  $h \rightarrow 0$ , we find that the birth rate of  $L$  in state  $i$  is  $\lambda$ . Similarly, the death rate of  $L$  in state  $i$  is  $\alpha\mu i/(i+k)$ . These are precisely the birth and death rates of the modified model with  $k$  special jobs. The equilibrium distribution of  $L$  is given by

$$\pi_{k+1}(i) = \frac{(i+k)!}{i!k!} \rho^i (1-\rho)^{k+1} , \quad i = 0, 1, \dots . \quad (12)$$

When  $k = 1$ , (12) coincides with (2).

Since  $L(t)$  represents the number of normal jobs present in the system at time  $t$ , the amount of service that a special job receives during the interval  $[0, A(x)]$  is equal to

$$\int_0^{A(x)} \frac{1}{L(s) + k} ds = \int_0^{A(x)} \frac{1}{B(C(s)) + k} ds = \int_0^x \frac{A'(u)}{B(u) + k} du = x . \quad (13)$$

In other words,  $A(x)$  is the time necessary for a special job to obtain service  $x$ .

Now consider the embedded Markov chain  $B(t_n)$ , where the sequence  $t_n$  forms a Poisson process with rate  $\mu$ . By the PASTA property, the equilibrium distribution of  $B(t_n)$  is given by (8). Hence, the equilibrium distribution of the embedded Markov chain  $L(A(t_n)) = B(t_n)$  is given by (8). But, according to (13), the sequence  $A(t_n)$  represents the consecutive instants when the special job completes exponentially distributed services with mean  $1/\mu$ , i.e., the consecutive feedback instants for a special job. Therefore, the equilibrium distribution of the number of normal jobs that a special job sees at feedback instants is (8), while the distribution seen by random observers and normal arrivals is (12), q.e.d.

## References

- [1] S.A. Grishechkin, “On the use of branching processes for finding steady state distributions in queueing theory”, *SIAM Journal on Theory of Probability and Applications*, 36, pp 477-493, 1992.
- [2] F.P. Kelly, *Reversibility and Stochastic Networks*, Wiley, 1979.
- [3] S.S. Lam and A.U. Shankar, “A Derivation of Response Time Distributions for a Multi-Class Feedback Queueing System”, *Performance Evaluation*, 1, pp 48-61, 1981.
- [4] S.S. Lavenberg and M. Reiser, “Stationary State Probabilities at Arrival Instants for Closed Queueing Networks with Multiple Types of Customers”, *J. Appl. Prob.*, 17, pp 1048-1061, 1980.
- [5] B. Melamed and D.D. Yao, “The ASTA Property”, Chapter 7 in *Advances in Queueing* (ed. J.H. Dshalalow), CRC Press, 1995.
- [6] K.C. Sevcik and I. Mitrani, “The Distribution of Queueing Network States at Input and Output Instants”, *JACM*, 28, 2, pp 358-371, 1981.



---

Unité de recherche INRIA Rocquencourt  
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)  
Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)  
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)  
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)  
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)  
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399